Decomposition of Data Matrices by Factors

- explore low dimensional structure in high dimensional space
- approximate the data matrix by a space of lower dimension
- solution(s) of optimization problem yields singular value decomposition of data
- the decomposition can be displayed graphically

Two ways approximation

- Two different ways to look at data matrix: $C(\mathbf{X})$ vs $C(\mathbf{X}^T)$
- Duality (very close connection) between the two views
- Application: image analysis, face recognition, fMRI, call center data, internet traffic data, etc...

Row vector space approximation

Rank 1 approximation

- Obtain the best 1-dimensional subspace that approximates the rows of $\mathbf{X}$

Row vector space approximation

Rank 1 approximation

- Obtain the best 1-dimensional subspace that approximates the rows of $\mathbf{X}$
- Find the straight line through the origin that approximates the rows of $\mathbf{X}$

Row vector space approximation

Rank 1 approximation

- Obtain the best 1-dimensional subspace that approximates the rows of $\mathbf{X}$
- Find the straight line through the origin that approximates the rows of $\mathbf{X}$
- Find $\mathbf{v} \in \mathbb{R}^p$ with $\mathbf{v}^T\mathbf{v} = 1$ such that $\sum_{i=1}^n ||\mathbf{x}_i - P_{\mathbf{v}}\mathbf{x}_i||^2$ is minimized

Row vector space approximation

Rank 1 approximation

- Obtain the best 1-dimensional subspace that approximates the rows of $\mathbf{X}$
- Find the straight line through the origin that approximates the rows of $\mathbf{X}$
- Find $\mathbf{v} \in \mathbb{R}^p$ with $\mathbf{v}^T\mathbf{v} = 1$ such that $\sum_{i=1}^n ||\mathbf{x}_i - P_{\mathbf{v}}\mathbf{x}_i||^2$ is minimized

  Note that $\sum_{i=1}^n ||\mathbf{x}_i - P_{\mathbf{v}}\mathbf{x}_i||^2 = \sum_{i=1}^n ||\mathbf{x}_i||^2 - ||P_{\mathbf{v}}\mathbf{x}_i||^2$

Row vector space approximation

Rank 1 approximation

- Obtain the best 1-dimensional subspace that approximates the rows of $\mathbf{X}$
- Find the straight line through the origin that approximates the rows of $\mathbf{X}$
- Find $\mathbf{v} \in \mathbb{R}^p$ with $\mathbf{v}^T\mathbf{v} = 1$ such that $\sum_{i=1}^n ||\mathbf{x}_i - P_\mathbf{v}\mathbf{x}_i||^2$ is minimized

  Note that $\sum_{i=1}^n ||\mathbf{x}_i - P_\mathbf{v}\mathbf{x}_i||^2 = \sum_{i=1}^n ||\mathbf{x}_i||^2 - ||P_\mathbf{v}\mathbf{x}_i||^2$
- Find $\mathbf{v} \in \mathbb{R}^p$ with $\mathbf{v}^T\mathbf{v} = 1$ such that $\sum_{i=1}^n ||P_\mathbf{v}\mathbf{x}_i||^2$ is maximized.

Row vector space approximation

Rank 1 approximation

- Obtain the best 1-dimensional subspace that approximates the rows of $\mathbf{X}$
- Find the straight line through the origin that approximates the rows of $\mathbf{X}$
- Find $\mathbf{v} \in \mathbb{R}^p$ with $\mathbf{v}^T\mathbf{v} = 1$ such that $\sum_{i=1}^n ||\mathbf{x}_i - P_{\mathbf{v}}\mathbf{x}_i||^2$ is minimized

  Note that $\sum_{i=1}^n ||\mathbf{x}_i - P_{\mathbf{v}}\mathbf{x}_i||^2 = \sum_{i=1}^n ||\mathbf{x}_i||^2 - ||P_{\mathbf{v}}\mathbf{x}_i||^2$
- Find $\mathbf{v} \in \mathbb{R}^p$ with $\mathbf{v}^T\mathbf{v} = 1$ such that $\sum_{i=1}^n ||P_{\mathbf{v}}\mathbf{x}_i||^2$ is maximized.

$$\begin{aligned}
\sum_{i=1}^n ||P_{\mathbf{v}}\mathbf{x}_i||^2 &= \sum_{i=1}^n (\mathbf{x}_i^T\mathbf{v})^2 \\
&= \sum_{i=1}^n \mathbf{v}^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{v} \\
&= \mathbf{v}^T(\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T)\mathbf{v} \\
&= \mathbf{v}^T\mathbf{X}^T\mathbf{X}\mathbf{v}
\end{aligned}$$

Row vector space approximation

Rank 1 approximation

- Obtain the best 1-dimensional subspace that approximates the rows of $\mathbf{X}$
- Find the straight line through the origin that approximates the rows of $\mathbf{X}$
- Find $\mathbf{v} \in \mathbb{R}^p$ with $\mathbf{v}^T\mathbf{v} = 1$ such that $\sum_{i=1}^n ||\mathbf{x}_i - P_{\mathbf{v}}\mathbf{x}_i||^2$ is minimized

  Note that $\sum_{i=1}^n ||\mathbf{x}_i - P_{\mathbf{v}}\mathbf{x}_i||^2 = \sum_{i=1}^n ||\mathbf{x}_i||^2 - ||P_{\mathbf{v}}\mathbf{x}_i||^2$
- Find $\mathbf{v} \in \mathbb{R}^p$ with $\mathbf{v}^T\mathbf{v} = 1$ such that $\sum_{i=1}^n ||P_{\mathbf{v}}\mathbf{x}_i||^2$ is maximized.

$$
\begin{aligned}
\sum_{i=1}^n ||P_{\mathbf{v}}\mathbf{x}_i||^2 &= \sum_{i=1}^n (\mathbf{x}_i^T\mathbf{v})^2 \\
&= \sum_{i=1}^n \mathbf{v}^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{v} \\
&= \mathbf{v}^T(\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T)\mathbf{v} \\
&= \mathbf{v}^T\mathbf{X}^T\mathbf{X}\mathbf{v}
\end{aligned}
$$

- Find $\mathbf{v} \in \mathbb{R}^p$ with $\mathbf{v}^T\mathbf{v} = 1$ such that $\mathbf{v}^T\mathbf{X}^T\mathbf{X}\mathbf{v}$ is maximized

Quadratic form maximization and the eigenvalues/vectors

Theorem: Let $\mathbf{A}$ be symmetric. Then,

$$\begin{aligned}
\max_{\mathbf{v}\in\mathbb{R}^p} \frac{\mathbf{v}^T\mathbf{A}\mathbf{v}}{\mathbf{v}^T\mathbf{v}} &= \max_{\mathbf{v}\in\mathbb{R}^p, \mathbf{v}^T\mathbf{v}=1} \mathbf{v}^T\mathbf{A}\mathbf{v} \\
&= \mathbf{v}_1^T\mathbf{A}\mathbf{v}_1 \\
&= \lambda_1,
\end{aligned}$$

where $(\lambda_1, \mathbf{v}_1)$ is the largest eigenvalue and associated normalized eigenvector of $\mathbf{A}$.

proof: Note that

$$\max_{\mathbf{v}\in\mathbb{R}^p} \frac{\mathbf{v}^T\mathbf{A}\mathbf{v}}{\mathbf{v}^T\mathbf{v}} = \max_{\mathbf{v}\in\mathbb{R}^p, \mathbf{v}^T\mathbf{v}=1} \mathbf{v}^T\mathbf{A}\mathbf{v}.$$

Let $\mathbf{A} = \sum_{i=1}^p \lambda_i \mathbf{v}_i \mathbf{v}^T = \mathbf{V}\Lambda\mathbf{V}^T$ be the spectral decomposition of $\mathbf{A}$, where $\lambda_1 \geq \cdots \geq \lambda_p$ are the eigenvalues of $\mathbf{A}$.

Then,

$$\begin{aligned}
\mathbf{v}^T\mathbf{A}\mathbf{v} &= \mathbf{v}^T\mathbf{V}\Lambda\mathbf{V}^T\mathbf{v} \\
&= \mathbf{w}^T\Lambda\mathbf{w}, \text{ where } \mathbf{w} = \mathbf{V}^T\mathbf{v}
\end{aligned}$$

and

$$\mathbf{w}^T\mathbf{w} = (\mathbf{V}^T\mathbf{v})^T(\mathbf{V}^T\mathbf{v}) = \mathbf{v}^T\mathbf{V}\mathbf{V}^T\mathbf{v} = \mathbf{v}^T\mathbf{v} = 1.$$

Quadratic form maximization and the eigenvalues/vectors

Thus,

$$\max_{\mathbf{v}\in\mathbb{R}^p, \mathbf{v}^T\mathbf{v}=1} \mathbf{v}^T\mathbf{A}\mathbf{v} = \max_{\mathbf{w}\in\mathbb{R}^p, \mathbf{w}^T\mathbf{w}=1} \mathbf{w}^T\Lambda\mathbf{w}.$$

Since $\mathbf{w}^T\Lambda\mathbf{w} \leq \lambda_1(\mathbf{w}^T\mathbf{w}) = \lambda_1$ and the maximum value, λ_1, is attained when

$$\mathbf{w} = (1, 0, \cdots, 0)^T$$

or equivalently

$$\mathbf{v} = \mathbf{V}\mathbf{w} = \mathbf{v}_1.$$

Row vector space approximation

Rank 1 approximation

- Find $\mathbf{v} \in \mathbb{R}^p$ with $\mathbf{v}^T\mathbf{v} = 1$ such that $\mathbf{v}^T\mathbf{X}^T\mathbf{X}\mathbf{v}$ is maximized

Row vector space approximation

Rank 1 approximation

- Find $\mathbf{v} \in \mathbb{R}^p$ with $\mathbf{v}^T\mathbf{v} = 1$ such that $\mathbf{v}^T\mathbf{X}^T\mathbf{X}\mathbf{v}$ is maximized
- Solution of the maximization problem is

$$\mathbf{v} = \mathbf{v}_1,$$

the first normalized eigenvector of $\mathbf{X}^T\mathbf{X}$ and the maximum value is given by λ_1, the largest eigenvalue.

- Representation of data on the 1-dimensional subspace:

$$\mathbf{x}_1^T\mathbf{v}_1, \;\; \mathbf{x}_2^T\mathbf{v}_1, \;\; \cdots , \;\; \mathbf{x}_n^T\mathbf{v}_1$$

Row vector space approximation

Rank 2 approximation?

- Find 2-dimensional subspace $F = \{c_1\mathbf{v}_1 + c_2\mathbf{v}_2\}$.
- Find $\mathbf{v} \in \mathbb{R}^p$ with $\mathbf{v}^T\mathbf{v} = 1$ and $\mathbf{v}^T\mathbf{v}_1 = 0$ such that $\mathbf{v}^T\mathbf{X}^T\mathbf{X}\mathbf{v}$ is maximized.

Quadratic form maximization and the eigenvalues/vectors

Theorem: Let $\mathbf{A}$ be symmetric. Then,

$$\begin{aligned}
\max_{\substack{\mathbf{v} \in \mathbb{R}^p, \mathbf{v}^T\mathbf{v} = 1 \\ \mathbf{v}^T\mathbf{v}_1 = 0, \ldots, \mathbf{v}^T\mathbf{v}_{k-1} = 0}} \mathbf{v}^T\mathbf{A}\mathbf{v} &= \mathbf{v}_k^T\mathbf{A}\mathbf{v}_k \\
&= \lambda_k,
\end{aligned}$$

where $\lambda_1 \geq \cdots \geq \lambda_p$ are the eigenvalues of $\mathbf{A}$ and $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are the associated normalized eigenvectors.

Row vector space approximation

Rank 2 approximation?

- Find 2-dimensional subspace $F = \{c_1\mathbf{v}_1 + c_2\mathbf{v}_2\}$.
- Find $\mathbf{v} \in \mathbb{R}^p$ with $\mathbf{v}^T\mathbf{v} = 1$ and $\mathbf{v}^T\mathbf{v}_1 = 0$ such that $\mathbf{v}^T\mathbf{X}^T\mathbf{X}\mathbf{v}$ is maximized.
- Solution? $\mathbf{v} = \mathbf{v}_2$ and the maximum value is λ_2.
- Representation of data on the 2-dimensional subspace:

$$(\mathbf{x}_1^T\mathbf{v}_1, \mathbf{x}_1^T\mathbf{v}_2),\ \ (\mathbf{x}_2^T\mathbf{v}_1, \mathbf{x}_2^T\mathbf{v}_2), \qquad (\mathbf{x}_n^T\mathbf{v}_1, \mathbf{x}_n^T\mathbf{v}_2)$$

Rank k approximation?

- Sequential maximization with additional orthogonal constraint.
- Solutions? $\mathbf{v}_1, \ldots, \mathbf{v}_r$ with the maximum values $\lambda_1, \ldots, \lambda_r$, where $r = \text{rank}(\mathbf{X})$.

Column vector space approximation

Rank 1 approximation

- Obtain the best 1-dimensional subspace that approximates the columns of $\mathbf{X}$
- Find the straight line through the origin that approximates the columns of $\mathbf{X}$
- Find $\mathbf{u} \in \mathbb{R}^n$ with $\mathbf{u}^T\mathbf{u} = 1$ such that $\sum_{i=1}^{p} ||\mathbf{x}_i - P_{\mathbf{u}}\mathbf{x}_i||^2$ is minimized where $\mathbf{x}_i$ denotes the i-th column of $\mathbf{X}$
- Find $\mathbf{u} \in \mathbb{R}^n$ with $\mathbf{u}^T\mathbf{u} = 1$ such that $\mathbf{u}^T\mathbf{X}\mathbf{X}^T\mathbf{u}$ is maximized.

Column vector space approximation

Rank 1 approximation?

- From the (quadratic form maximization) Theorem,

$$\max_{\mathbf{u}\in\mathbb{R}^n, \mathbf{u}^T\mathbf{u}=1} \mathbf{u}^T(\mathbf{X}\mathbf{X}^T)\mathbf{u} = \mathbf{u}_1^T(\mathbf{X}\mathbf{X}^T)\mathbf{u}_1 = \nu_1$$

  where ν_1 is the largest eigenvalue of $\mathbf{X}\mathbf{X}^T$ and $\mathbf{u}_1$ is the corresponding unit length eigenvector.

Column vector space approximation

Rank 1 approximation?

- From the (quadratic form maximization) Theorem,

$$\max_{\mathbf{u}\in\mathbb{R}^n, \mathbf{u}^T\mathbf{u}=1} \mathbf{u}^T(\mathbf{X}\mathbf{X}^T)\mathbf{u} = \mathbf{u}_1^T(\mathbf{X}\mathbf{X}^T)\mathbf{u}_1 = \nu_1$$

  where ν_1 is the largest eigenvalue of $\mathbf{X}\mathbf{X}^T$ and $\mathbf{u}_1$ is the corresponding unit length eigenvector.

Rank k approximation?

- Sequential maximization with additional orthogonal constraint.
- Solutions? $\mathbf{u}_1, \ldots, \mathbf{u}_r$ with the maximum values $\nu_1, \ldots, \nu_r$, where $r = \text{rank}(\mathbf{X})$.

Duality between two way approaches

Let $\{(\lambda_i, \mathbf{v}_i)\}$ and $\{(\nu_i, \mathbf{u}_i)\}$ be the eigenvalue-eigenvector pairs of $\mathbf{X}^T\mathbf{X}$ and $\mathbf{X}\mathbf{X}^T$, respectively. Then,

$$\mathbf{X}^T\mathbf{X}\mathbf{v}_i = \lambda_i \mathbf{v}_i.$$

Pre-multiplying by $\mathbf{X}$ on both sides, we get

$$\mathbf{X}\mathbf{X}^T\mathbf{X}\mathbf{v}_i = \lambda_i(\mathbf{X}\mathbf{v}_i)$$

or equivalently

$$\mathbf{X}\mathbf{X}^T(\mathbf{X}\mathbf{v}_i) = \lambda_i(\mathbf{X}\mathbf{v}_i).$$

This means that λ_is also also eigenvalues of $\mathbf{X}\mathbf{X}^T$ and $\mathbf{X}\mathbf{v}_i$ are eigenvectors of $\mathbf{X}\mathbf{X}^T$.

Since $\mathbf{u}_i$ are length-one vectors,

$$\mathbf{u}_i = \frac{\mathbf{X}\mathbf{v}_i}{L(\mathbf{X}\mathbf{v}_i)} = \frac{\mathbf{X}\mathbf{v}_i}{\sqrt{\mathbf{v}_i^T\mathbf{X}^T\mathbf{X}\mathbf{v}_i}} = \frac{\mathbf{X}\mathbf{v}_i}{\sqrt{\lambda_i}}.$$

Similarly,

$$\mathbf{v}_i = \frac{\mathbf{X}^T\mathbf{u}_i}{\sqrt{\lambda_i}}.$$

Once $\{(\lambda_i, \mathbf{v}_i)\}$ are computed, don't need to recompute $\{(\nu_i, \mathbf{u}_i)\}$.

Duality between two way approaches: SVD

Write this relationship,

$$\mathbf{u}_i = \frac{\mathbf{X}\mathbf{v}_i}{\sqrt{\lambda_i}}, i = 1, \ldots, r$$

in a matrix format:

$$[\mathbf{u}_1, \mathbf{u}_2 \ldots, \mathbf{u}_r] = \mathbf{X}[\mathbf{v}_1, \mathbf{v}_2 \ldots, \mathbf{v}_r] \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 & \ldots & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \frac{1}{\sqrt{\lambda_r}}. \end{pmatrix} \tag{1}$$

Define $\mathbf{U}_{n\times r} = [\mathbf{u}_1, \mathbf{u}_2 \ldots, \mathbf{u}_r]$, $\mathbf{V}_{p\times r} = [\mathbf{v}_1, \mathbf{v}_2 \ldots, \mathbf{v}_r]$, and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$, then (1) can be written as

$$\mathbf{U} = \mathbf{X}\mathbf{V}\Lambda^{-1/2}$$

or by post-multiplying $\Lambda^{1/2}\mathbf{V}^T$

$$\mathbf{X} = \mathbf{U}\Lambda^{1/2}\mathbf{V}^T.$$

The two-way low rank approximation provides Singular Value Decomposition (SVD) of data matrix.

SVD

Singular Value Decomposition (SVD) of data matrix

$$\mathbf{X}_{n\times p} = \mathbf{U}\Lambda^{1/2}\mathbf{V}^T$$

provides

- sequential
- solutions to the best approximation of data
- in the column and the row spaces
- simultaneously.

The two-way rank k approximation of the data is:

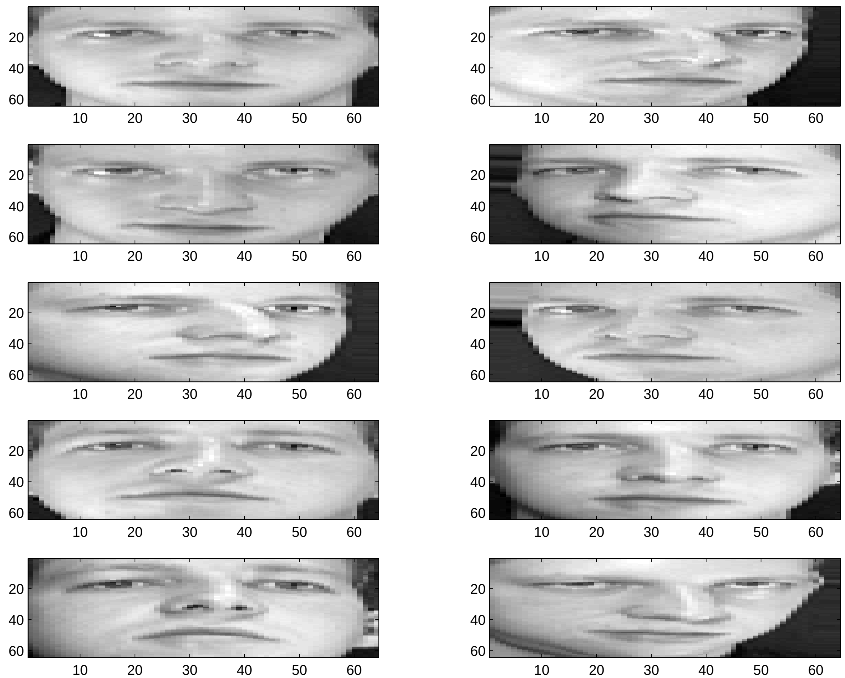
$$\mathbf{X}_{n\times p} \approx \mathbf{U}_k\Lambda_k^{1/2}\mathbf{V}_k^T,$$

where $\mathbf{U}_k = [\mathbf{u}_1, \ldots, \mathbf{u}_k]$, $\mathbf{V}_k = [\mathbf{v}_1, \mathbf{v}_2 \ldots, \mathbf{v}_r]$, and $\Lambda_k = \text{diag}(\lambda_1, \ldots, \lambda_k)$ for $k = 1, \ldots, r$.

Example: Face Image

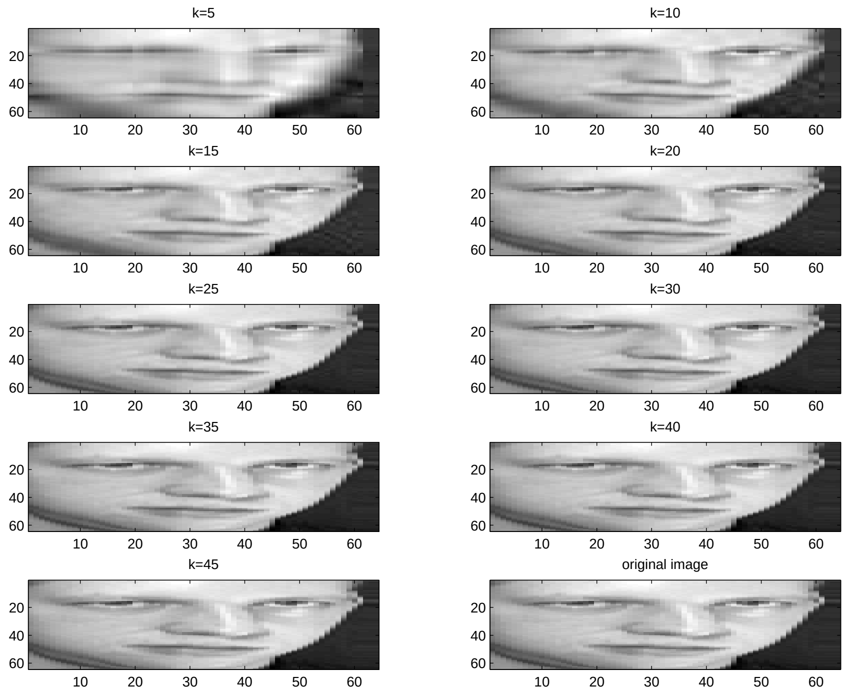
data: http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

Example: Face Image



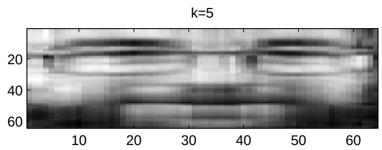
Take the last picture of size 64×64.

Example: Face Image approximation



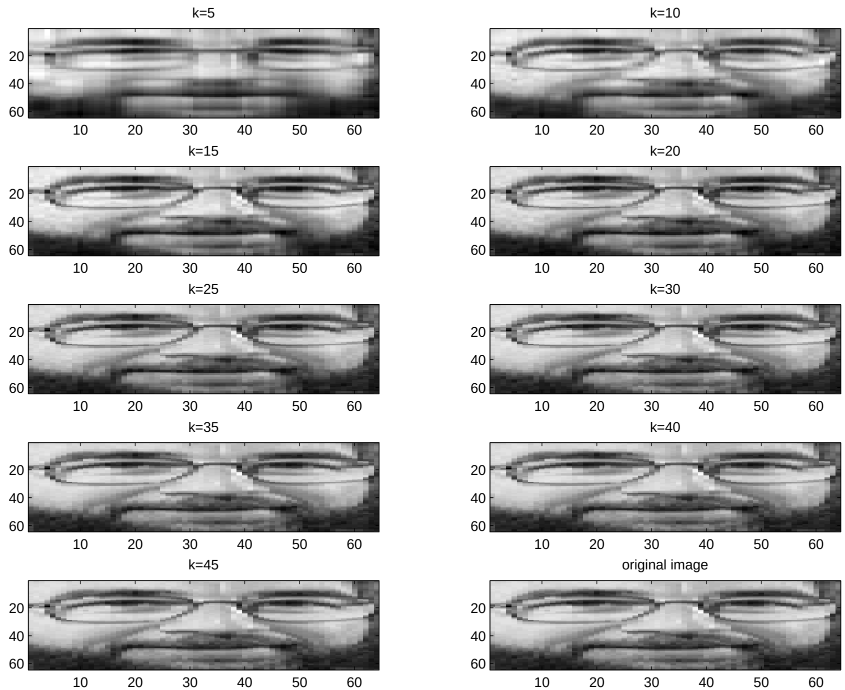
Rank k approximation using SVD.

Example: Face Image approximation

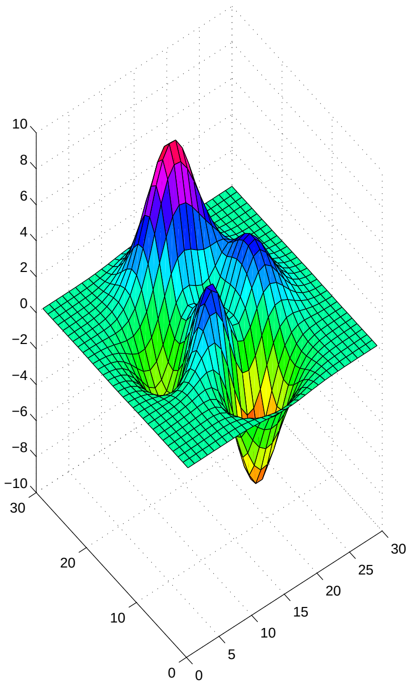


Rank $k = 5$ approximation using SVD.

Example: Face Image approximation

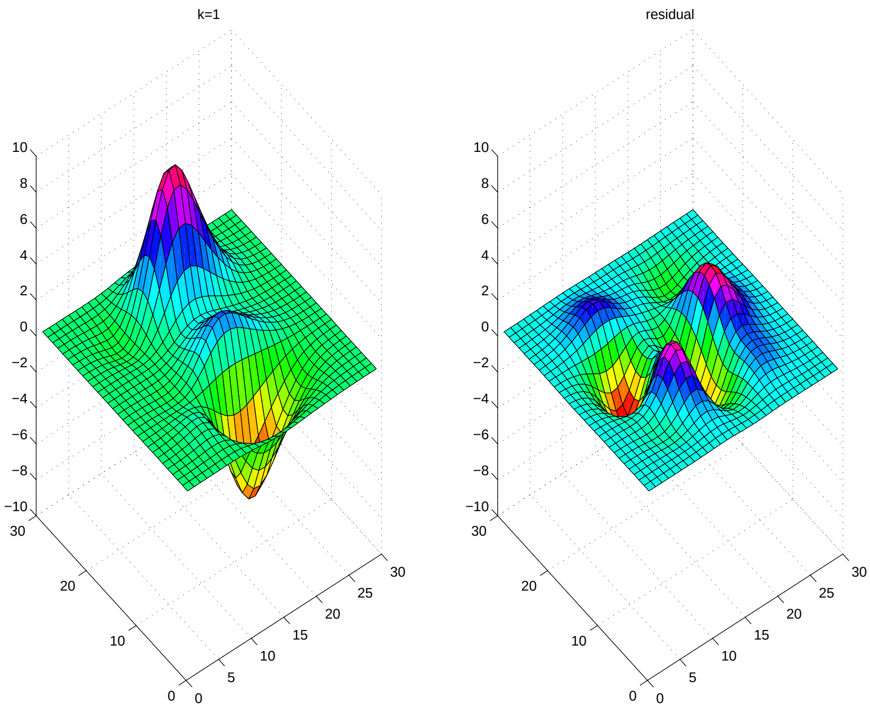


Example: Matrix in a surface plot



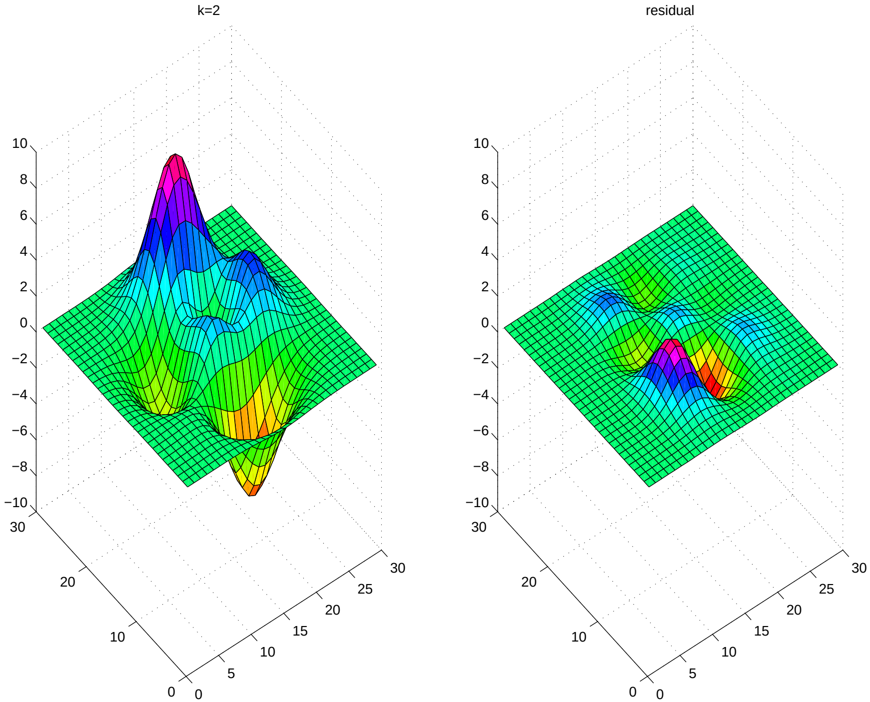
Data matrix of size 30×30.

Example: Matrix in a surface plot



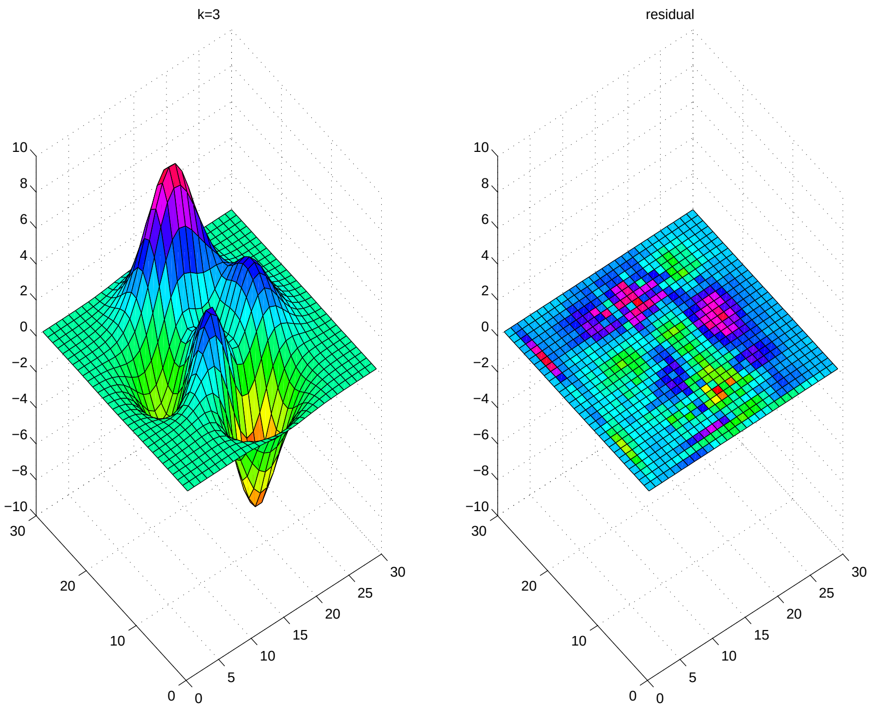
Rank $k = 1$ approximation using SVD.

Example: Matrix in a surface plot



Rank $k = 2$ approximation using SVD.

Example: Matrix in a surface plot



Rank $k = 3$ approximation using SVD.